Qualitative and Quantitative Methods in Libraries (QQML) 4: 709-–715, 2015

Natural language search in image collections

Béla Lóránt Kovács[1], Margit Takács[2]

[1]Affilation: University of Debrecen, Faculty of Informatics
[2]Affilation: University of Debrecen, Faculty of Informatics

Abstract: It is a challenge for today's library practice to make the entities in digital images retrievable according to the users demands. In our presentation we wish to introduce a searching method which is able to seek certain elements and their environment in the images by applying a search within them in natural language. One key is to this method is extending the Dublin Core metadata system by a new element or by a classifier of the description element which help to enlist the natural language denominations of the elements in the images and the elements' positions within the image. Following this we are able to find certain elements then calculate and grade the information value of their environment with the help of a modified version of Shannon's entropy formula. The application of this method enables us to perform further value calculations for certain details within a particular image which make further, more efficient searches possible.
Key words: image collections, searching by image, metadata, Dublin Core, data structure, information, intelligence value, entropy

1. Introduction

Last decades with the appearance of electronic libraries there was a growing number of images besides texts in digital collections (Eakins 1999). In these databases it was rather limited to describe images and search them with traditional library methods (Goodrum 2000, Matusiak 2006). Altogether their data related to their titles, authors, creations and other conditions can be analysed by using traditional library methods. To describe their content and search them by natural language phrases is possible only in a rather complicated way (Yee et al. 2003). Therefore in our presentation we look for an answer to this question: how it is possible to conduct a natural language search in image collections.

Received: 21.10.2014 / Accepted: 1.3.2015 ISSN 2241-1925

2. Metadatas

In order to make it possible, first we have to examine metadata systems used in libraries. Among them Dublin Core (DC) (ISO 2009) is the most widespread and well-known in library world. The first 13 elements of DC were created in 1995. They were intended for describing textual sources available on internet. The DC has been changing since that time, newer and newer elements have been added to it. The DC element set was complemented with qualifiers which specify the meaning of each element (e.g. the date element can be qualified with date of production, publication and modification). Due to this DC can be better used for the cataloguing of library collections also. Later on DC was extended to non-textual, mainly image documents.

From the more significant 15 elements ISO standard was constructed which has already appeared in several countries, among them in Hungary too. The standard entitled *Information and documentation – Dublin Core metadata element set* became a description standard of information sources which extended beyond special fields. The electronic documents are typical information sources for Dublin Core applications. The standard determines only that element set which in general is used in relation to some task or application. Requirements and aspects applied at a given place and in a certain community can make additional restrictions, regulations and interpretations necessary. It is not objective of the standard to determine detailed criteria which the element set is used with for special tasks or applications (ISO 2009).

A special application like this retrieves entities which can be seen in traditional and digital images according to user needs in some collections. Therefore we have to apply a search method which finds each element in the images and applies a natural language search in its context. For this purpose those metadata is needed to be used which list elements in the images with their natural language names and positions.

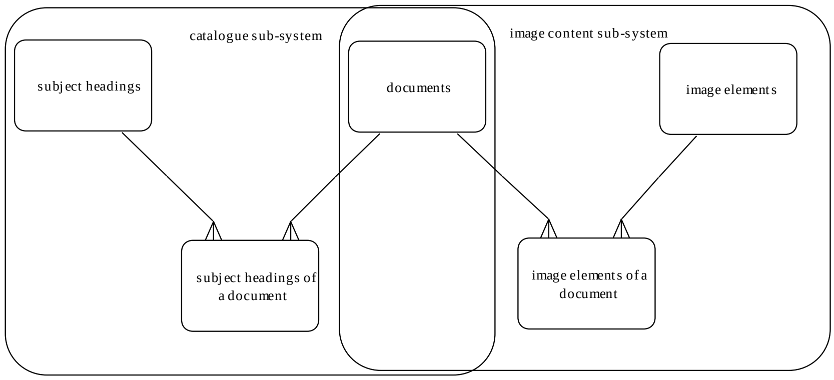
Concerning Dublin Core we have three possibilites for subject description. We can describe topic of the information source with keywords, subject headings and classification numbers by using Subject element. Description element provides subject description which can be an abstract, a citation, table of contents, a reference to a graphical presentation of the content or a freely formulated description of the content. Third possibility for presenting the content is the Coverage element that means content or field of application of information source in space and time (extension). In the case of digital images the elements which are available in the standard are suitable for classifying each image into the appropriate category and describing events and happenings in the images and placing the content of the image or the image itself in space and time. Although all the three data can mean a lot in finding an image, but none of them can be used for searching each element in the images or searching its context.

To list entities in a given image the Description element can be best used among from the three above mentioned metadata. It is true that with the help of this element we can provide a brief content descrition, but with the introduction of the 'elements' qualifier it is becoming suitable for listing entities in the image too. We can also give the position of certain entities by *'position'* qualifier. Due to the two qualifiers newly introduced the natural language search is already becoming possible. Before presenting search for image elements and their information value it is also necessary to define the more important concepts that we will use during our presentation.

3. Data structures

At data storage and retrieval we will apply the following concepts: image document, image collection, image sub-collection, subject heading and image element. The *image document* is a physical carrier of information which users can perceive as a view with appropriate technical devices. The *image collection* is a structured database which consists of image documents and their attributes (metadata). The image sub-collection is a part of image collection where image documents are available with at least one same subject heading. The *subject heading* is an attribute which is used for thematic grouping of image documents. The image element is a part of the image document. It can carry some kind of information for users and which we can supply with a natural language attribute. The newly introduced qualifiers of Dublin Core, i.e. the *dc.description.elements* and the *dc.description.position* can be applied for describing image elements and their positions. When we develop database structure of the search system, we will use these concepts.

We considered the relational data model (Codd 1969, Ullman 2008) to be the best one for the development of data structure of a search system. Our data model has three main components such as the catalogue sub-system, the image content sub-system and the image details information sub-system. We will obtain the *catalogue sub-system* with the conversion of OPAC data which includes the bibliographic descriptions of the documents. In the data structure of the sub-system the *subject headings* are very important. In data received from the OPAC the subject heading is a repeatable metadata. Since it is a significant attribute from a semantic aspect as well which after taking over need to be complemented if required. For this reason it is practical to place subject headings in a separate table. We place also the group code of the image sub-systems in this table. In the *image content sub-system* it is important to support giving a name to the image details and to provide authority control of names. Image details are connected directly to the documents table and they contain *dc.description.elements* or *dc.description.position* metadata which are repeatable attributes with regard to one document, for this reason we store them in a separate table. A librarian who is good at cataloguing has to supply image documents with subject headings or image elements with the appropriate metadata. We illustrate the relationship of the catalogue sub-system and the image content sub-system in the following figure:



Besides the above mentioned two sub-systems we need to complement the data structure with a third sub-system, i.e. the *image details information sub-system.* It is developed for search interface of the database. It contains index files for browsing and search. The most important services of the sub-system: browsing (in the most significant attributes of the document, in subject-headings and in image details), search (basic, advanced, according to information content and intelligence value). After all these we have to find the answer only to this question: on the basis of what formulas we can calculate information value of the image documents.

4. Formulas

To calculate the information value of image elements we use Norbert Wiener's formula (Wiener 1948). In this formula I indicates the information value of element x which we calculate in the following way:

$$I(x) = -\log_2 p(x). \qquad (1.)$$

Here p(x) means the probability of occurrence of element x within the image sub-system. We obtain the probability of occurrence of the image element in the sub-system by taking the ratio of all occurrences of all elements there (number of all cases) and all occurrences of element x (number of favoured cases) in the following way: $\frac{\text{number of favoured cases}}{\text{number of all cases}}$.

In an image sub-system we can give values to image documents and sort them with the help of two formulas. The first one computes *information value* of documents, the second one calculates the *intelligence value* of documents within the sub-system. However we wish to define the concept of intelligence value a bit later. We can receive the information value carried by image documents

I(ID) if we simply add together information value of elements in the image according to the following formula:

$$I(ID) = -\sum_{i=1}^{n} \log_2 p(x_i). \quad (2.)$$

The $p(x_i)$ is the probability of occurrence of one element in the image within the sub-system, n is the number of elements in the image. According to this, the image document will be identical with the elements which appear in it (ID = x_1, x_2 … x_n). During a search by means of this formula we can sort documents according to which one has the highest or the lowest information value in a collection.

We count *intelligence value* of the image documents with the following formula. We mean by intelligence value that on average how much information is carried by one element of the image. Intelligence value is often much more revealing than information value in the case of one document in a sub-system. It shows how new or usual the image is in a certain field. The intelligence value of an image document H(ID) can be calculated with the following formula:

$$H(ID) = -\sum_{i=1}^{n} \frac{1}{n} \log_2 p(x_i). \quad (3.)$$

The $p(x_i)$ is the probability of occurrence of one element in the image within the sub-system, n is the number of elements in the image. Measure of the intelligence value is bit/element. Its formula shows a remarkable similarity with Shannon's entropy formula (Shannon 1948), for this reason it received a traditional sign of the entropy, H.

In addition to the above mentioned we can also count what information value is associated with a certain image element within a given document of an image sub-system. Here we can calculate two things as well. In the first case we can compute how much information value the context of an image element has in relation to the element. The $I(Cx_1)$ is information value of context of the image element x_1 that we obtain with the following formula:

$$I(Cx_1) = -\sum_{i=2}^{n} \log_2 p(x_i). \quad (4.)$$

Here the $p(x_i)$ is probability of occurrence of the element x_i in context of element x_1 within the sub-system, n is the number of elements in the image. We obtain the $p(x_i)$ value in the sub-system by taking the ratio of all occurrences of all the other elements in the context of x_1 element (number of all cases) and all occurrences of element x_i (number of favoured cases) in the following way:

$\frac{number\ of\ favoured\ cases}{number\ of\ all\ cases}$. (Of course element x_1 can occur several times as well in an image, so it also has a probability that this element emerges twice, three times or several times in an image. More times it appears less probability it has. Therefore we have to introduce $p(x_{1.1})$, $p(x_{1.2})$ … $p(x_{1.z})$ signs which indicate the probability of the repeated occurrences within the image context).

In the second case we compute that how much intelligence value a context of an image element has, i.e. on average how much information is carried by the other image elements in connection with it. The $H(Cx_1)$ is intelligence value of context of the image element x_1 that we count in the following way:

$$H(Cx_1) = -\sum_{i=2}^{n} \frac{1}{n-1} \log_2 p(x_i). \qquad (5.)$$

Here the $p(x_i)$ is probability of occurrence of the element x_i in context of element x_1 within the sub-system, n is the number of elements in the image. From our aspect perhaps this value is the most interesting because it shows how new or usual an image or an image detail is from the aspect of the given element. If we conduct a search for a phrase then we can sort documents in the sub-system according to the intelligence value of context of an image element specified by the query. The document with the highest value will reveal the most news about the image element, while the one with the lowest value will depict it in the most usual way.

The database management system can operate formulas with its own functions. When a new document enters the database, database management system will recalculate various probabilities or information values of occurrence of each image element every time with its functions. The user can make use of these values for different aims. On the search interface with the application of these values he can e.g. browse and search in index files. Not only a user but a graphical software can also use these values. The task of this software can be to make calculations with vector values related to the image elements. We stored these vectors by means of *dc.description.position* metadata in the image content sub-system of our database. The software counts what probabilities an image element can occur with at one position of one document in an image sub-collection. It can also calculate for us what information values an image element can have at a given position. Information values arising from the frequency of occurrences and image positions can be added together. All this can make possible further search modes for the users, but on a later occasion we wish to give a talk on it.

5. Conclusions

On the basis of the above mentioned we can conclude that it is possible to conduct a natural language search in image collections of digital libraries. First it was necessary to construct the appropriate metadata system that we achieved with modification of the Dublin Core (ISO 2009). Secondly an appropriate data structure was required that we accomplished with the use of a relational data model (Codd 1969, Ullman 2008). Finally we had to determine those formulas which the functions of a search system used for sorting image documents by their information values or intelligence values. These functions are based on mathematical theory model of communication, first of all on Norbert Wiener (Wiener 1948) and Claude E. Shannon's research (Shannon 1948). During a search we had to introduce a new concept of intelligence value. It indicated average information content of all the other elements which appeared in context of an element in one of the images, or it meant average information content of all the other elements which appeared in context of all elements in one image. The measure of intelligence value is bit/element. Each image element was specified with natural language phrases, so we could use these phrases during a search. The search method can be applied in the most diverse image collections and after further developments – on the basis of our current research - it can be connected with the search systems of full-text databases as well.

References

E.F. Codd, (1969), Derivability, Redundancy, and Consistency of Relations Stored in Large Data Banks, *ACM SIGMOD Record*, Volume 38 Issue 1, 17-36.

John Eakins – Margaret Graham, (1999), Content-based Image Retrieval, *JISC Technology Applications*, 4.

Abby A. Goodrum, (2000), Image Information Retrieval: An Overview of Current Research, Informing Science, 3, 63-66.

ISO 15836, (2009), Information and documentation - The Dublin Core metadata element set

Krystyna K. Matusiak, (2006), Towards user-centered indexing in digital image collections, *OCLC Systems & Services*, 22, 283-298.

Jeffry D. Ullman – Jenifer Widom, (2008), *A Firs Course in Database System*, Pearson Education Inc, Pearson Prentice Hall

Claude E. Shannon, (1948), A Mathematical Theory of Communication, *The Bell System Technical Journal*, 27, 379-423. 623-656.

Norbert Wiener, (1948), *Cybernetics: or Control and Communication in the Animal and the Machine*, The MIT Press, Cambridge

Ka-Ping Yee – Kirsten Swearingen – Kevin Li – Marti Hearst , (2003), Faceted metadata for image search and browsing, *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, New York, 401-408.